

Common structural variation reduction in PLS-SEM: Replacement analytic composites and the one fourth rule

Ned Kock

Texas A&M International University, USA

Abstract

Path coefficients may be distorted, in the context of structural equation modeling via partial least squares (PLS-SEM), due to excess common structural variation shared in a model. This may be caused by methodological issues; e.g., the use of highly correlated but conceptually distinct latent variables, or common method bias. We discuss a common structural variation reduction procedure using WarpPLS, a leading PLS-SEM software tool. This procedure relies on the creation of analytic composites as replacements for latent variables, where the weights are one fourth of the original path coefficients among the latent variables and their predictors in the structural model, and with signs that are the opposites of the signs of the original path coefficients.

Keywords: Collinearity; Common Method Bias; Analytic Composites; Path Coefficients; Structural Equation Modeling; Partial Least Squares; WarpPLS.

Introduction

The amount of common structural variation shared in a model, in the context of structural equation modeling via partial least squares (PLS-SEM), can be directly measured through the full collinearity variance inflation factors (FCVIFs) calculated for the latent variables (LVs) in the model (Kock, 2015; Kock & Lynn, 2012). The more common variation is shared among the LVs in the model, the higher are the FCVIFs.

Whenever the highest FCVIF in a model is above a predefined threshold (e.g., 3.3, 5 or 10), path coefficients may be distorted due to the excess common structural variation shared in the model. This may occur due to a number of reasons, such as the use in the model of highly correlated LVs that nevertheless measure different constructs, highly correlated indicators for different LVs, or significant common method bias (Kock, 2015; Kock & Lynn, 2012).

In these cases, PLS-SEM users may want to reduce the amount of common structural variation shared in their models. This can be done with WarpPLS, a leading software tool that implements classic composite-based and more modern factor-based PLS-SEM algorithms (Kock, 2019a; 2019b), among other convenient features (Amora, 2021; Hubona & Belkhamza, 2021; Kock, 2020b; 2020c; 2020d; 2021; Morrow & Conger, 2021; Moqbel et al., 2020).

We discuss a common structural variation reduction procedure in this article. This procedure relies on the creation of analytic composites as replacements for LVs. The weights used to create the analytic composites are one fourth of the original path coefficients among the LVs and their

predictors in the structural model (i.e., their predictor LVs), and with signs that are the opposites of the signs of the path coefficients. For simplicity, and without any impact on the generality of our discussion, all variables are assumed to be standardized.

Common structural variation reduction: The one fourth rule

Let X be an LV that is caused by two other LVs, A and B , according to the equation below, where ε_X is the structural error term that accounts for the variance in X that is not explained by A or B . This three-LV configuration provides the basis for a general discussion that applies to more complex models.

$$X = \beta_{XA}A + \beta_{XB}B + \varepsilon_X.$$

This leads to X sharing a certain amount of variation with A and B , measured by the correlations r_{XA} and r_{XB} . By performing the operation specified in the assignment equation below, where w_{XA} and w_{XB} are weights that are generally expected to be lower than r_{XA} and r_{XB} , we create an analytic composite that aggregates X , A and B .

$$C = \text{Stdz}(X - w_{XA}A - w_{XB}B).$$

In this assignment equation $\text{Stdz}(\cdot)$ is the standardization function. Here C can be seen as version of X in which the amount variation that comes from A and B is reduced. We call this operation *common structural variation reduction* because it reduces the amount of variation shared by the LVs that make up a structural model, which together with the measurement model (linking LVs and their indicators), form the basis for an SEM analysis.

This common variation reduction operation can be employed to produce variables C and X that are correlated strongly enough to be seen as redundant, but that are not identical in that C incorporates less variation from A and B . Two variables are redundant if the VIF calculated based on their scores alone is greater than 10, which is associated with a correlation of 0.950 or higher.

If the highest FCVIF in such a model is above a predefined threshold (e.g., 3.3, 5 or 10), what are good starting values for weights to generate a variable C so that the highest FCVIF in such a model is reduced? Based on a number of simulations, as well as discussions with WarpPLS users based on their empirical studies, it appears that good starting values for the weights w_{XA} and w_{XB} are one fourth of the corresponding path coefficients β_{XA} and β_{XB} .

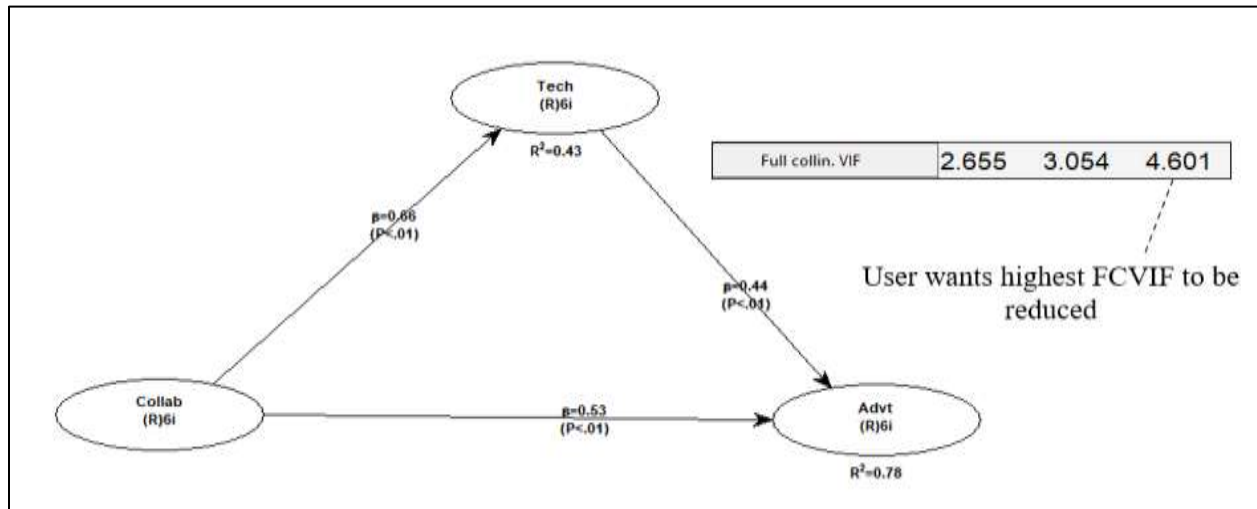
Using analytic composites to reduce FCVIFs

The illustrative model shown in Figure 1 is used as a basis for our discussion. A similar model has been used elsewhere (see, e.g., Kock, 2015). As it can be seen, it contains three LVs. We created 300 rows of data employing the Monte Carlo simulation method (Kock, 2016). Also shown are the FCVIFs calculated for our model. These are measures of common variation among the LVs in the model. The highest FCVIF stands at 4.601. Let us assume that a researcher wants to reduce this highest FCVIF.

Since the FCVIFs shown refer to LVs, and not to indicators, the first step to reduce the highest FCVIF is to create a new model where the LVs are measured through their scores. To do this, we use the menu option “Add all latent variable (a.k.a. factor) scores as new standardized indicators” to add the LV scores to the dataset as new standardized indicators. In Figure 2, these indicators are named “lv_Tech”, “lv_Collab” and “lv_Advt”. After these indicators are added to

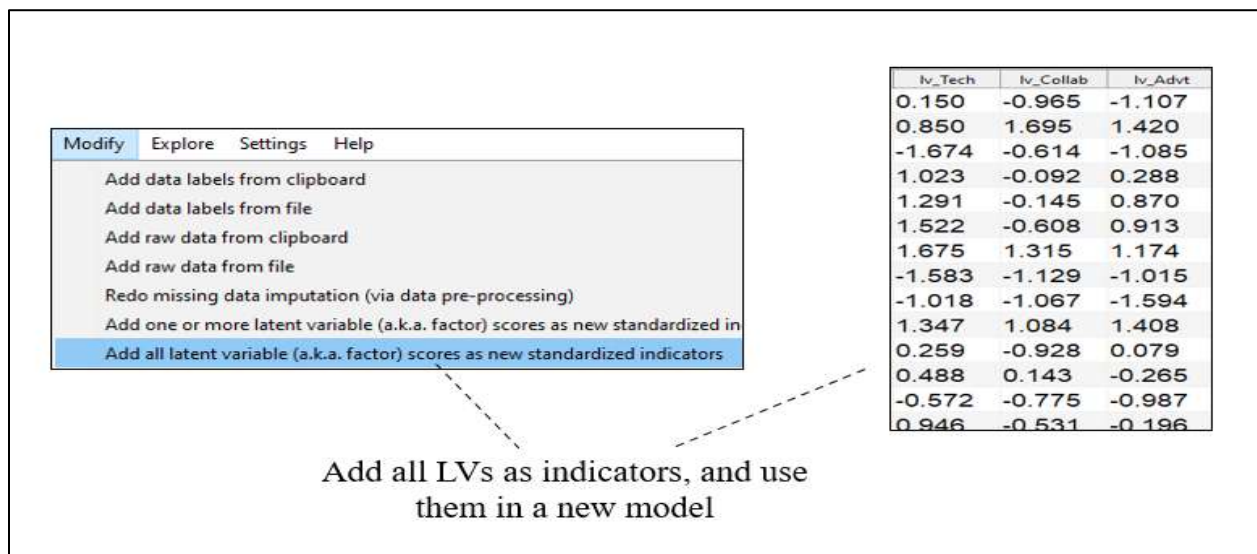
the dataset, we then use them to measure the LVs. The resulting model contains only single-indicator LVs.

Figure 1: Illustrative model used



Notes: Collab = collaborative culture; Tech = e-collaboration technology use; Advt = competitive advantage; notation under latent variable acronym describes measurement approach and number of indicators, e.g., (R)6i = reflective measurement with 6 indicators.

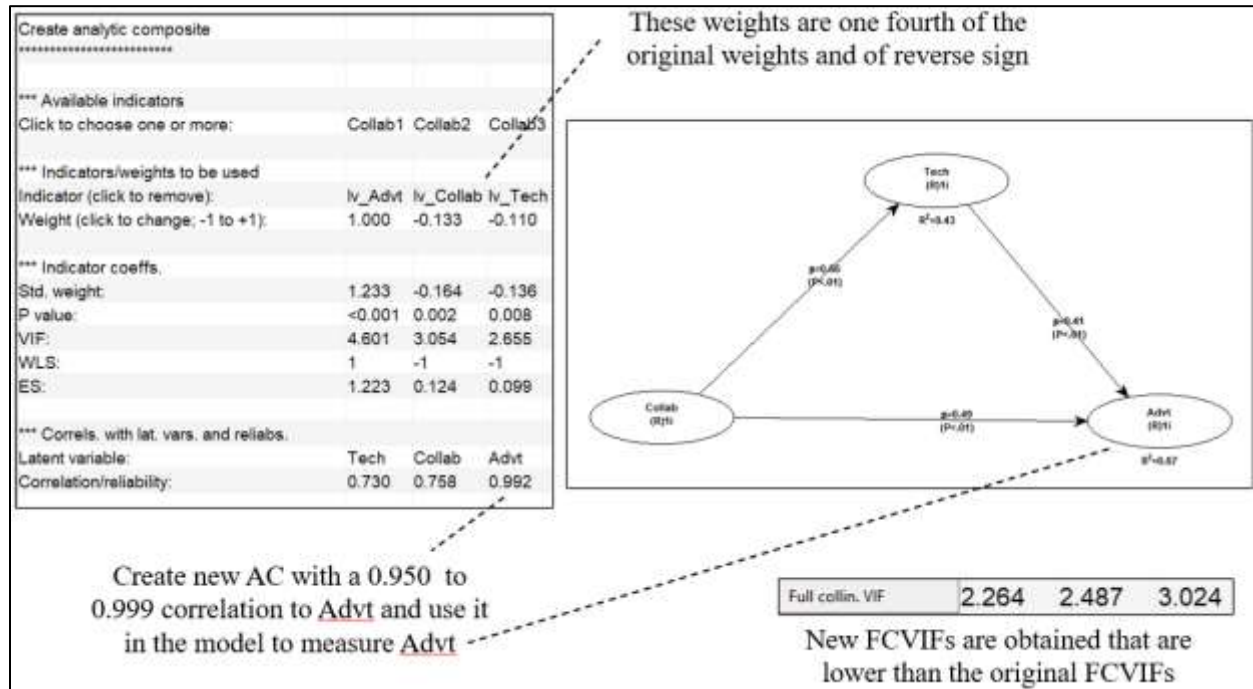
Figure 2: Using LV scores in a new model



The next step is to create an analytic composite that aggregates “lv_Tech”, “lv_Collab” and “lv_Advt” according to user-defined weights. The weights used for “lv_Tech” and “lv_Collab” are one fourth of the original path coefficients among these two LVs and “lv_Advt”, and with signs that are the opposites of the signs of the path coefficients.

This is shown in Figure 3, where the weights used for “lv_Tech” and “lv_Collab” are -0.133 and -0.110 respectively. These are calculated by dividing the original path coefficients of 0.530 and 0.440 by 4 and multiplying them by -1. Since the analytic composite is expected to be used in place of “lv_Advt” in the model, the weight used for “lv_Advt” is positive 1.

Figure 3: Creating and using analytic composite



For the analytic composite to be redundant with “lv_Advt”, which is desirable, the combination of weights used should be so that the correlation between the analytic composite and “lv_Advt” is in the 0.950 to 0.999 range. The former is the lowest correlation at which these two variables are redundant at the VIF > 10 level. Note that the correlation is 0.992 in our example.

Once the analytic composite is created, we use it in place of “lv_Advt” in the model, as the sole measure of “Advt”. Next, we conduct Step 5 again, obtaining new results. Among these new results are new FCVIFs, which are now lower than the FCVIFs in the original model. As we can see, the new FCVIFs for “Collab”, “Tech” and “Advt” are, respectively: 2.264, 2.487 and 3.024. These are lower than the original 2.655, 3.054 and 4.601.

Through this procedure we created an analytic composite whereby we removed variation in “Advt” that had come from “Collab” and “Tech”, with the goal of reducing the FCVIF of 4.601 associated with “Advt”. Yet all FCVIFs went down in value. The reason for this is that each FCVIF is calculated taking all model LVs into consideration.

Conclusion

Note that we removed variation in “Advt” from “Collab” and “Tech” in a way that is consistent with our model, and thus with the theory being tested through the model, since in the

model “Collab” and “Tech” are causes of “Advt”. When an LV causes another in a model, or is hypothesized to influence another, the underlying theory assumes that it in fact transfers variation to the LV it causes or influences.

Removing variation in a way that is consistent with the underlying theory is arguably preferable to doing it so otherwise, even though the final FCVIFs may go down in the same general way. For example, if we had removed variation in “Collab” using “Advt” and “Tech”, the FCVIFs would have gone down as well, but the final results would be different. If our procedure does *not* lead to FCVIF reductions, this could be an indication that the model is incorrectly specified.

Also note that the weights we used for “lv_Tech” and “lv_Collab” to create the analytic composite, namely -0.133 and -0.110, led to FCVIFs for “Collab”, “Tech” and “Advt” that could be seen as lower than desired. If we had used weights of lower absolute magnitude, we would have obtained greater FCVIFs. For example, we could have used, at first, weights that were one eighth of the original path coefficients, as opposed to one fourth.

Our common structural variation reduction procedure had a strong impact on FCVIFs, even though we aimed for correlations between the analytic composite and “lv_Advt” that were clearly indicative of redundancy. Given this, researchers using this technique may target correlations at the high end of the 0.950 – 0.999 range at first, by using weaker weights, and then progressively move toward 0.950, stopping when the highest FCVIF reaches a level that is seen as acceptable.

Acknowledgments

The author is the developer of the software WarpPLS. He is grateful to WarpPLS users for questions, comments, discussions, and continued use.

References

- Amora, J. T. (2021). Convergent validity assessment in PLS-SEM: A loadings-driven approach. *Data Analysis Perspectives Journal*, 2(3), 1-6.
- Hubona, G., & Belkhamza, Z. (2021). Testing a moderated mediation in PLS-SEM: A full latent growth approach. *Data Analysis Perspectives Journal*, 2(4), 1-5.
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1-10.
- Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, 15(1), 299-315.
- Kock, N. (2019a). From composites to factors: Bridging the gap between PLS and covariance-based structural equation modeling. *Information Systems Journal*, 29(3), 674-706.
- Kock, N. (2019b). Factor-based structural equation modeling with WarpPLS. *Australasian Marketing Journal*, 27(1), 57-63.
- Kock, N. (2020a). *WarpPLS User Manual: Version 7.0*. Laredo, TX: ScriptWarp Systems.
- Kock, N. (2020b). Full latent growth and its use in PLS-SEM: Testing moderating relationships. *Data Analysis Perspectives Journal*, 1(1), 1-5.
- Kock, N. (2020c). Multilevel analyses in PLS-SEM: An anchor-factorial with variation diffusion approach. *Data Analysis Perspectives Journal*, 1(2), 1-6.

- Kock, N. (2020d). Using indicator correlation fit indices in PLS-SEM: Selecting the algorithm with the best fit. *Data Analysis Perspectives Journal*, 1(4), 1-4.
- Kock, N. (2021). Harman's single factor test in PLS-SEM: Checking for common method bias. *Data Analysis Perspectives Journal*, 2(2), 1-6.
- Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.
- Moqbel, M., Guduru, R., & Harun, A. (2020). Testing mediation via indirect effects in PLS-SEM: A social networking site illustration. *Data Analysis Perspectives Journal*, 1(3), 1-6.
- Morrow, D. L., & Conger, S. (2021). Assessing reciprocal relationships in PLS-SEM: An illustration based on a job crafting study. *Data Analysis Perspectives Journal*, 2(1), 1-5.